# THE DETERMINATION OF
# EMPIRICAL AND ANALYTICAL SPACECRAFT PARAMETRIC CURVES
# - THEORY AND METHODS -

PROGRESS REPORT I

Submitted

by the

TEXAS A&M RESEARCH FOUNDATION

to the

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

for NASA Grant

SC-NGR-44-001-027

Prepared

by

THE INDUSTRIAL ENGINEERING DEPARTMENT

of

TEXAS A&M UNIVERSITY

December 1, 1965

# TABLE OF CONTENTS

# F O R E W O R D

This document represents the first progress report on the NASA research grant NGR 44-001-027. The report is divided into five general areas which may be considered individually or collectively. A summary of these sections is presented below as an outline of the contents of the report and the information that is contained therein.

PART I is the introduction and contains an overview of how the individual parts of the research are being directed toward a common objective and how these results are oriented toward the specific problems of the granting agency.

PART II contains a philosophy of cost model development which discusses some of the possibilities open in the development of cost models such as those presently being considered by the Long Range Planning Office at the Manned Spacecraft Center.

PART III summarizes another area of research which is oriented toward the specific analytical problem of determining parameters associated with elements of a mathematical model by using nonlinear estimation of those parameters in a restricted space.

PART IV summarizes a third research area which is directed toward the collection, analysis, and utilization of subjectively determined data in the construction and application of mathematical models.

PART V contains the advisory memos that have been written as part of the services performed under this grant. These include critiques of other NASA contractor reports as well as informal meetings held at NASA and Texas A&M.

PART VI is a report on a completed research project in the area of learning curves which is considered as appropriate to this study since one of the basic elements in cost model work considers the effects of learning on the costs.

PART VII contains a Bibliography of the current literature in the general areas of the research that is being conducted as part of this grant.

## PART I - INTRODUCTION

The specific problem that is the basis for this grant and the direction of the research that is being performed is that of fitting functions to data in the generation of functional relations which describe the cost of subsystems in spacecraft and the testing of those functions to determine their credibility with the respect to the model in which they are incorporated. Toward this end there is presently three areas of research being conducted as part of the grant. These three areas are discussed individually in Parts II, III, and IV. If it is not obvious from considering each of the areas individually, this section of the report is an effort to tie these parts into a integrated approach to the problem.

In the research discussed in Part II, the research is directed at the end result - the final model, the appropriate composition of that model, and deriving the functions which describe the response to various levels of the variables. The second area of research which supports this is the Part IV in that it is concerned with providing the functions which are used in the developments in Part III. However, this part of the research is concerned primarily with the attainment of consistency and statistical credibility in subjective data which must necessarily be used due to little or no data being available for use in deriving the function. Therefore, Part IV of the research will provide the automated means whereby data will be collected and analyzed prior to providing the functions required in Part II.

There is still a futher backup effort that is an area of research and is that discussed in Part III of the report. This research essentially supports Part IV by developing the statistical and mathematical background for fitting functions to a set of data points. This is in order to provide some statistical properties of the functions which are fitted to the collected data. This is a more academic type of research but is necessary in order to develop the basic theory which will permit the model development to progress on a sounder statistical basis. Therefore, the data obtained and analyzed in Part IV will eventually use the methods of fitting functions in a constrained parameter space that are developed in Part III of the research.

The general benefits that will accrue to the granting agency will primarily come directly from Part II of the research, however it is strongly supported by that work which is being performed in the other two areas of research. The work that is described in Part V keeps the study group apprised of the current problems in this research area and will permit the results of the work performed under this grant to be directed more specifically to current problems, rather than abstracted in some form which must go through a transition phase prior to application to the immediate problems in the cost model development area.

The work exemplified in Part VI and VII is a continuing effort to provide backup, conduct investigations and provide information in basic areas that can best be done by a university type of research group. This will provide the support in depth that may be used to integrate and understand the range of the problems that have been treated previously. The applicability of independent research in these basic areas is to assure full coverage of the techniques and possible misapplications of techniques that are introduced into the planning and cost model areas.

In summary it is felt that the simultaneous direction of the research that is outlined and reported on in this document is in agreement with the proposed research and in accordance with the direction given by the granting agency. It is anticipated that this grant will provide some original approaches to old problems in the cost modeling area and the use of cost models in the planning approach.

# P A R T II

This section of the report contains some to the general philosophy of cost model development which has been discussed informally with Long Range Planning personnel at NASA/MSC at various times. The discussions of this section are generally based upon other NASA work and some other applicable model building experience. Some cost model evolution is traced in this report to permit the model builder to apply the particular technique in the hierarchy presented that best suits the amout of information that he has available and the time limitation he may encounter. This does not summarize completely the research that can be done in this area as it is a continuing area for research and will be a continuing part of the research grant task.

# A PHILOSOPHY OF COST MODEL DEVELOPMENT

When in the course of long range planning, it becomes necessary to develop a cost model there are still some methods which have not been tested for the collection of data and formulation of the model. This paper will be directed toward the discussion of model formulation outside of the tired methods which attempt to duplicate statistical rigor with two to five data points. This paper presents some of the concepts which will be developed further during the course of this research grant.

A limited amount of historical data appears to be a recurring problem or obstacle in the development of meaningful cost models for futuristic type designs. This problem is not unique to either cost models or spacecraft and it is encountered in planning of development type or advanced technology in both cost, time, reliability and performance areas. Therefore, the general subject is one of the lack of data and its influence upon building mathematical models. The developments in this paper will primarily be from the applied standpoint and should be easily followed with a minimum amount of mathematics.

The primary application of this research appears to be in the cost model area in the near future, specific reference will be made where applicable to this type of work. Many of the concepts which are being discussed below have been communicated to the sponsors of this grant during verbal discussions at various meetings. The basic idea is that the experts and the planners in this field have information which is in a qualitative form at present; however, if it were converted to a quantitative form, it could provide the basic mechanism for the cost estimation procedures which are now used. In addition it would provide the added advantage, in that the numbers that are provided in the estimation process could be reconstructed at some time in the future to provide even more historical information for the model building process and the error estimation process as repeated from time to time in the model building phase of cost models.

One of the basic ideas is that there are cost dominating variables generally contained in cost models. For example, during World War II, sufficiently accurate cost models could be produced through the analysis of the weight of the airplane. This essentially meant that these planes were of the same general technological level and this was an accurate predictor due to the averaging out which took place through the total development and production effort. However, spacecraft development and production has some different aspects in that it is composed primarily of a research and development type of effort and the production runs can not be counted upon for the effects mentioned above. It is also possible that the accounting and control methods presently used monitor projects to the point that the averaging that previously took place is eliminated.

However, due to the technologies that are encountered in the various parts of the spacecraft, it becomes more obvious that there are significant cost influences present in the cost other than the weight. This becomes more obvious from the engineering standpoint by a simple analysis of the composition of a subsystem.

This is futher supported by the long list of causal variables which were produced by the Booz-Allen study presently in progress. Generally it is agreeded among those persons consulted in this area that any one of the variables involved in the set of causal variables could conceivably become a cost dominating influence to the point that the single variable could be used as the single independent variable in a cost equation. It is just as readily agreeded that there is an interaction among variables which contribute to the cost of a subsystem. However, this interaction is controlled to some extent by the design constraints on the variables.

Since this is a future planning type cost research project, it is felt that it should not be limited by the conventions of the past and that the idea of developing methods for handling this type of model should be the primary point of consideration. If we are to progress in this area some latitude must be taken, some experimentation permitted and some setbacks suffered in order to make incremental gains in this type of model building. Therefore, the following is an approach to model building which has been applied in a modified form in the area of rating of aircraft by pilots with very good results. The major deviation in this research is the extending of the concepts to account for interaction and the utilization of subjectively determined data in the development of the relationships. Part IV of this report discusses the research being carried out in the area of the collection, analysis and assimilation of these types of data. In this section it will be assumed that a single estimate in the form of a continuous function over the range of the variable involved will be available to the model builder and will be accepted as representing the function described by the model builder.

This procedure will normally be approached by the use of a single function used to relate the response of the variable which is the dependent variable of the model for the other independent causal variables in the model. In this case the response will be measured in terms of the cost of the subsystem or a larger division of the vehicle being considered. The single variable will be any one of the causal variables which is considered to influence cost.

There is a common statistical procedure which has been used on other occasions and may be used in this type of situation if some estimate of the variance is available for each of the predictors of the cost and if these are statistically independent estimating functions for the cost of the subsystem. If this is the case, these estimators may be combined by weighting them inversely proportional to their variances. This is a rather common procedure in mathematical statistics in the derivation of statistical models of physical processes; however, this has not been commonly applied previously in cost models but is a simple direct method of deriving a single cost equation with a larger number of independent variables with a minimum of effort. The statistical assumptions must be observed and the model builder should be aware of the implications if they are ignored. This technique was suggested approximately two years ago in another cost model study[1] and the statistical proof is presented by Graybill.[2] It is assumed that some estimate of the variance associated with a functional equation will be obtained in the data that is developed by the methods discussed in Part IV.

It is generally considered that this is an approximation method at best and it is generally agreed that there is some statistical interaction of the estimators in that they are not independent and that they are functionally dependent. Therefore, this method has become mature and it is possible to use this as a base and to build the methodology for combining estimates based upon the information that is available to the model builder. This presents the problem of evaluating the physical interaction as to what it means in terms of the cost and to develop some method of evaluating the statistical interaction or assessing the role of the physical interaction as it relates to the cost of the subsystem. In this area, the research has produced some results in describing the form of the interaction by a geometric model for the cost interaction. The procedure will be first, to rank the variables in the order of the most important from the standpoint of influencing the cost of the subsystem. This is not absolutely necessary in the early developments; however, the reasons for this will become more apparent in the evaluation of the interaction in later stages of the study. If one were to rank weight as the single most important variable in the functional relationship which describes the cost of the subsystem, it would then be assumed that some range of values would be available on all the other variables contained in the equation for that same cost. Only in the instance of one of the other variables becoming cost dominant would it be necessary to recognize its function in the causal cost relationship. However, as it has been pointed out in the above discussion, spacecraft are more subsecptible to minor variation in the parameters than are state-of-the-art aircraft.

The next step in the procedure is to consider the next most important variable from the standpoint of cost for consideration as to its role in the functional equation. For the purposes of this example, assume that this variable is the volume of the spacecraft. This function is or may be some function of the weight also, in that density may be used to define the dependent relationship between the volume and the weight, however, the expression for the cost contributed by the volume of the spacecraft should be costed an increment above that which is considered as being contributed by the weight. Therefore, the volume will provide the incremintal cost above that being accounted for by the weight of the spacecraft. Now consider the problem associated with evaluating just this part of the model and how it would work. First of all, it appears important to point out that there are a number of levels of sophistication that could be entered into at this point; however, the simplest method will be discussed in an attempt to promote understanding and establish the groundwork for more sophisticated methods which will be explained during the grant period.

Assume that through the methods of Part IV it has been possible to determine the form of the cost to weight relationship as shown in Figure 1 for a given type of subsystem. This would be

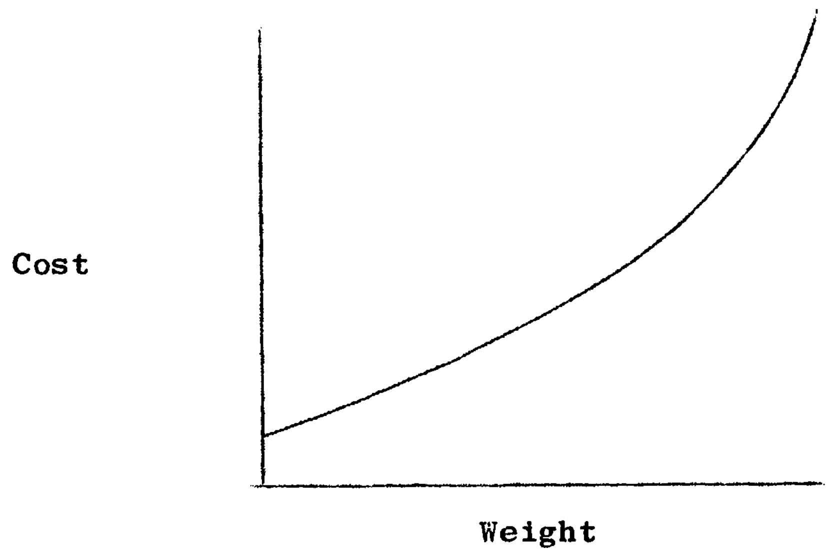



Figure 1. Cost as a function of the weight of a specific type of subsystem.

used if only a single variable were to be used in the cost estimation relationship. The function illustrated represents the cost of the subsystem not only as a function of weight, but it will represent a certain nominal level of other variables that describe the subsystem. In other words, this is not simply the cost of a passive subsystem which only has weight, it has volume, peice parts, operating pressure, velocity and all the variables that can be used to describe it and which could conceivably be cost related. It is assumed that the function shown above also represents the cost of the functional interaction (to be distinguished from statistical interaction) with all the other variables associated with the system. Now consider the inclusion of another variable in the cost functional relationship. This variable will be entered such that it complements the variable already included in the equation and does not duplicate the costing function represented by the weight. This second variable for the purposes of an example will be volume. Lets assume that a function similar to that illustrated in Figure 1 can be determined for the volume of the subsystem. This presents the problem of combining these two functions into a single cost estimating fuction. Obviously, to describe interaction among two variables it is necessary to use a three dimensional space to discuss the interaction. If the trace of the cost response surface on the cost - weight plane and the cost - volume plane is considered, one would obtain the two independent predictor functions for their respective cost dominant case. However, to be realistic, it would be necessary to consider these to be inclusive of all the other causal variables contained in the cost function represented by the trace on the plane. In each case it would be considered that the level of the other variable is a nonsignificant quantity rather than having a value of zero. If the surface which represents the cost of a spacecraft of a given wieght and volume is represented for all points in the three dimensional space, it represents the relation among all three variables and the interaction of the two variables is properly represented by the response surface. It would be possible to generate this response surface by using a brute force approach that would require the response of one variable to be estimated over its range for fixed incremental values of the other variable over its range. This would be a rather laborious process even for two variables. It is easily seen that this would be a prohibitive method if the variables numbered three or four for a single relationship and each variable had a nominal range.

In order to use the subjective data collection approach to provide the desired information for the model construction it is necessary to reduce the number of functions to be estimated by the expert. This is the present status of the research in that an iso-cost cutting plane is being used to determine the amount of information that can be provided by a limited number of subjectively determined functions. Essentially the procedure is one of using the constant cost cutting plane to intersect the surface of the cost response that is a function of two variables. The trace of the surface on the cutting plane provides a starting point for the development of methods for the collection of data. This trace is bounded on each end by the cost dominant condition of the corresponding variable. The methods that are being considered for generation of the traces on the cutting planes and subsequently the entire response surface are based upon the sensitivity of cost to changes in one variable at various levels of the other variable. The contribution of each variable to the total cost may possibly be estimated in terms of correlation coefficients. This approach presently appears to reduce the amount of data necessary by a significant amount; however, some of the questions that need to be answered by the experts will require a good understanding of the mechanics of the technique being used. The present effort is in simplifying the data collection for obtaining estimates of the interaction.

It becomes obvious at this point that the other two research areas outlined in this report are integrated within the problems of the development of a cost model building philosophy. In addition to the problem described above, there has been a geometric model constructed for three variables and the corresponding cost response. These are illustrated by the use of a tetrahedron where three of the surfaces are the individual cost-variable relationships and the fourth surface contains iso-cost contours to describe the cost response associated with the three variables. The physical interaction is established by the relationship of the individual functions which are represented by lines in three dimensional space. Even though the geometric interpretation and demonstration of this model is relatively simple to describe, the mathematics involved in converting it to a cost model is a little more difficult. As soon as this is accomplished, it will probably provide the key to expanding to higher order interactions and more complex models.

The research conducted thus far in the study has led to a number of paths that are being explored simultaneously. These include the study of a dimensionally based cost function which could simplify the subjective data procedure. If this proves to be a feasible approach, the work being done in the fitting of functions in a constrained parameter space will permit progress at a rapid rate. The use of partial derivatives is also being considered in simplifying the model construction. The use of partial correlation coefficients in the derivation of the responses was mentioned briefly above, the specific problem is obtaining good subjective estimates of those values. The overall plan is to develop a modeling procedure that will provide additive elements which represent each new variable that contributes to the cost. In order to do this efficiently, the model will probably contain a number of delta functions which will introduce various elements of the overall relationship as a function of the level of the respective variables.

1. Launch Vehicle Systems Cost Model , FZM - 4070, Progress Report 2, 27 February, 1964, General Dynamics / Fort Worth.

2. Graybill, F. A., An Introduction to Linear Statistical Models, McGraw-Hill, New York, 1961, page 409.

# PART III

The research outlined in this section of the report is directed toward the estimation of the parameters of a mathematical model representation of some process where the parameters themselves are constrained. The assumptions associated with the form of the model are very general and are directed specifically to the type of problems being encountered in the construction of cost models. This research is an effort to extend the theoretical work that should be offered in support of model development and the estimation of parameters. Much of the preliminary research has been acomplished as indicated by the statement of the problem and in the bibliography which is contained in PART VII. The general approach and the tools for the solution of the problem have been hypothesized and presently the stage of the investigation is the formulation of the fitting problem as a programming problem. The next area of research will be oriented more specifically to the statistical properties of the estimators of the parameters.

The information contained in this section is necessarily limited, due to the fact that the total development of the solution is not available at this time and any report on the direction of the present investigation would be premature and might necessarily be refuted at some later report date; however, this section presents the general approach and the problem definition.

# ESTIMATION OF PARAMETERS WITH CONSTRAINTS FOR MODELS WHICH ARE NONLINEAR IN THE PARAMETERS

The general objectives of this research are as follows: (1) To estimate the parameter vector $\Theta$ in the function $y = f(X, \Theta)$, where $X = (x_1, x_2, \ldots, x_p)$ is the vector of observable variables, $\Theta = (\theta_1, \theta_2, \ldots, \theta_k)$ is a vector of fixed but unknown parameters, y is observable, (depending on X), f is non-linear in the parameters $\Theta_i$, and $\Theta$ is constrained to a certain region of $E^k$ (k-dimensional Euclidean Space). (2) To determine the statistical properties of $\hat{\Theta}$ (the estimator for $\Theta$) with a view to determining $\hat{\Theta}$ so that it has various properties that are "good" in a statistical sense, such as unbiasedness, minimum mean-square error and consistency. Both "large sample" and "small sample" properties of $\hat{\Theta}$ will be investigated, with the emphasis on the "small sample" case-i.e.-what one does when a minimum of data is available. (3) To establish condidence regions for $\Theta$.

It is anticipated, in veiw of the results obtained by applying linear programming procedures to linear problems of this type, that the general field of mathematical programming may yield a solution of the non-linear problem. Convex programming will be of particular interest since it is expected that most of the constraints will be convex, and also that least squares estimation (which involves a convex function) will be utilized.

Various results from response surface methodology will be studied (since the least squares problem may be approached as a response surface problem) with the hope that an estension of some of these methods will throw light on the problem of what to do about the constraints.

Minimum absolute deviation (rather than least squares) will also be investigated as a possible criterion for "best fit".

Various modes of attack on the problem were discussed above. However, since this is an unsolved problem in general, it is not known which, if any, of these methods will yield a solution. It may be conjectured, on the basis of experience, that a "marriage" of response surface and mathematical programming techniques will yield at least a numerical solution for certain classes of constrained non-linear problems. It may also be assumed that no single technique will be "optimum" for <u>all</u> constrained non-linear problems.

A simple function of the type that will be under study is $y = f(X, \theta_1, \theta_2, \theta_3, \theta_4) = \theta_1 e^{\theta_2 X} + \theta_3 e^{\theta_4 X}$.

If (1) the parameters are unconstrained, if (2) the X's are equally spaced, and if (3) the number of data points are exactly 4, a "fit" for this equation is known (which may or may not be "best").

However, if <u>either</u> (1), (2), or (3) does not hold, a solution to estimating the parameters $(\theta_1, \theta_2, \theta_3, \theta_4)$ in some "optimum" fashion is not readily available.

It is apparent that work in the area of non-linear estimation is needed, since no solution is now available for even this simple problem.

It is anticipated that the complete answer to the questions raised in the above will be obtained as a result of this research.

If, however, only some of them are answered, a large addition to the existing knowledge about estimation under uncertainty will have been made. The theory of linear estimation is well established at present, representing years of work by hundreds of investigators. Non-linear estimation has been relatively ignored, not so much because of a lack of application as because of the difficulty of the problem. Perhaps it can be shown that the problem is not unsolvable; needing only a fresh approach to spur renewed interest in it.

# PART IV

This section reports on one concept presently under development that will provide for the collection, assimilation and analysis of subjectively determined data in the construction and formulation of mathematical models. This will consider the statistical aspects, which include the bias, consistency, statistics of extremes and convergence for these types of data. This area of research has been designated in a effort to extend model building capability beyond the limitations frequently encountered due to the lack of historical data. An integral part of the study will be to develop methods for sampling subjective data, fitting prescribed functional forms and providing for an adaptive response in the model to the addition of actual data.

UTILIZATION OF SUBJECTIVELY DETERMINED DATA IN THE FORMULATION OF MATHEMATICAL MODELS

Model building that is designed to assist in futuristic type event descriptions often encounters a very serious limitation in terms of the approach that is generally considered the appropriate method of developing the mathematical forms that are included within the model structure. It is generally expected that the model developments will be based upon a form of fitting some type of mathematical function to existing data. The method will then generally use selected data points to check the capability of the mathematical function as a descriptor of the process that is being modeled. Estimates of the error or the variance of the estimators can then be obtained from the existing data in an unbiased manner. It is also well established by statisticians that the relationships derived are only applicable over the range of the variables that were used in the derivation of the relationship. There are two major problems which eminate from this type of philosophy in mathematical model development, they are: (1) in the spacecraft area there is very little historical data available on which to base the cost of future vehicles to be used in space exploration due to the technological changes that are being experienced in spacecraft development and the variance of the subsystems themselves; and (2) in planning for development type efforts there is generally an extrapolation beyond the range of present historical data in which case the applicability of the fitted function is questionable, this is especially true if there is a change in the technology required or as the variables involved in the function increase. These two restrictions have caused a great deal of distortion of statistical techniques and cause the approaches to development of models to be less than credible. Therefore, it is felt that it is reasonable to explore a more appealing and rational approach to the be used in obtaining information for use in the models that are being used for management and planning decisions of futuristic process that differs significantly from the past experience.

One of the primary reasons for the development and use of models in management decision making is due to the dimensionality of the problem involved and the mind has trouble comprehending all influences of various decisions or aspects of the problem. Therefore, if a subjective approach is to be used in the development of mathematical models through regular means, it is generally necessary to isolate each of the contributing factors and then to combine them is some logical fashion such that in their combined form they tend to represent the process which is being modeled. In any type of data collection process, it becomes the job of the model builder to collect that data which is required in order to isolate these main effects and their interaction such that the model can be constructed of basic elements which are available to the model user, and which can be used to input the information required by the model. The inputs required along with the model structure which employs the mathematical relationship constitutes the total job of the model builder.

It is relatively easy to obtain subjective opinion ; however, the problem is the agreement among several subjective opinions and the systematic processing of that data to obtain meaningful results from such data. With this type of requirement, it is then necessary to determine the best methods and the use of those data which may be obtained through the use of expert opinion. Since the expert can only consider a limited number of factors at a single time, it is necessary to develop techniques which will be capable of taking estimates of a limited number of factors at a single time and through the collection of a number of estimates, combine these estimates in such a fashion that they represent the total process for which the model is being built

This approach of using subjectively determined data will explore the collection of subjective data and the proper method for combining that data in a model formulation. This latter area of combining the various factors will be applicable to the normal methods of formulation, because in Part II it was indicated that the interaction is a weak area of the present mathematical model building techniques. Therefore, it will be necessary to formulate methods for determining the degree and direction of the interaction from the standpoint of the response of the model.

The techniqures that are developed in this research will be applicable to model building in general, but will be applied in the cost model area as the data becomes available. The general approach will be to develop questionnaires which relate to the particular area being studied and the factors which are felt to affect that area. It is anticipated in the cost area that all the pertinent variables have been defined by the other contracted cost work that is being done for NASA/MSC, and that the experience gained in these studies will contribute to the initial task of questionnaire development for collection of the data. Based upon previous experience in this area it is expected that the collection of this type of data and its analysis will be facilitated by the development of a computer program to assist in the statistical analysis and the computation of functional relationships. Some of the ideas advanced in Part II of this report will be used in the actual formulation of this model; however, the concepts developed through this research will help in collection of the necessary information for persuing the research as outlined above.

There is some indication that methods similar to that proposed in this section have been tried elsewhere, however, in those cases where they have been applied, it appears that certain statistical procedures could have improved the analysis. It is assumed that the attitude in this case has been one of, since it is a guess anyway, why bother with any degree of statistical rigor. However, in some previous work which cannot be quoted in this report, it was shown that statistical test of the multiple range type indicated an inability to distinguish between some questions which were directed toward assessing a subjective value. Upon brief interrogation of a sample of respondents to this particular questionnaire, this observation was confirmed; however, the results of the survey were published in their original form without using the statistical tests as pertinent information.

In summary, this part of the sponsored research is expected to result in the development of a general computer model which can be used in the collection, testing and utilization of subjectively determined data up through the first order interaction effects. If the research progresses smoothly, the development of the means for collection of higher order interaction information will be approached. The level of sophistication will be primarily a function of turnaround time which is provided by the experts contributing to this study. This model will take into consideration the full range of consistency, and the statistical aspects of the techniques applied to those data which are quantified as a result of this procedure. Presently a demonstration cost model which uses subjectively determined data is being placed on the computer system at Texas A&M preliminary to experimentation which is required in this developmental effort.

# P A R T V

This section of the progress report represents the informal consultation that is provided as a part of the cost research grant. The memoranda which are contained in this section are more than records of the meetings and should be reviewed for their technical content as it relates to the specific topic being considered by the individual memo.

To: NASA/MSC/LRP

From: Texas A&M/Industrial Engineering Department

Subject: Review of Manned Spacecraft Cost Analysis Program, Second Oral Briefing and Progress Report, September 28, 1965 at MSC Houston, Texas

This progress report presentation was attended as part of the service provided by the Texas A&M Research Foundation under NASA research grant NGR 44-001-027. This memo is provided as a general critique of the presentation and is in lieu of extended participation in the general discussion which followed the presentation.

The A&M position is one in which to be of the most benefit to NASA and MSC it must be aware of the actual progress being made in the related cost studies. Therefore, based upon this presentation, it is necessary that a number of pointed questions be asked in order to coordinate the research effort at Texas A&M and the contracted work being performed for MSC.

Since the oral presentation was closely parallel to the documentation contained in the brochure, this critique will be related to specific points within the brochure, rather than rely on recall of specific oral statements.

Page 2 Para. 2. How do the second and third objectives differ from what has already been determined by other cost studies contracted by NASA? (General Dynamics/ Fort Worth; Lockheed, Burbank; Lockheed, Sunnyvale; Rand).

Page 2 Para. 3. Can Mercury data be used as a check point by summing up the detailed costs of other programs to correspond to the level available in the Mercury data?

Page 2 Para 4. Later in this report, reference is made to the use of engineering judgement and weight is contained in the list of technical variables being considered for the CER's.

Page 4 Para 1. Are the present cost category definitions at a meaningful level for obtaining costs? That is, have they been successful in breaking out costs accurately into each of the cost categories defined and are they meaningful from the standpoint of being easily identified.

Page 4 Para 2. Since this is the mid tern report, it would be beneficial to the A&M cost research grant, if some of the methodology that they are developing were used on the data available in order to illustrate the problems and permit us to offer some guidance in the selection of techniques.

Page 12 Para 3. How are variables and combinations being tested?

Page 14 Para 1. Is the evaluation of interactions a suggestion or a statement of achievement, if so how is it accomplished and how is the state of the art judged? (Refer back to comment on Page 2 Paragraph 4.)

Page 18 Para 2. This paragraph tends to indicate that time dependency of various costs relative to other costs are being considered as part of this study.

Page 20 Para 3. Wasn't part of the study to provide a consistent breakdown of the costs, if only collection of data were required that would probably have already been accomplished?

Page 22 Para 4. How do they test the fit of the curve in the Work Flow Chart? Isn't it simply a matter of picking a curve?

Page 24. Are only linear progress functions available? Has learning been significant on any spacecraft programs? What has this work contributed? Where are production rate effects reflected?

Page 26 Para 3. Can the definition of "first" be based on similarity of production methods without consideration of time between units of production.

Page 36 Para 3. Will the total technical analysis be documented in order that individual CERs may be traced back through the analysis and reviewed for possible change? What are some of the criteria for a variable to be included in the analysis?

Page 40. Are the plots shown on a linear scale? Why weren't all the data points listed on Page 39 shown on Page 41?

Page 42. If the statements of this page are taken literally by assuming 10 different variables and only 4 trial exponent values, then the number of trials would be $\sum_{x=2}^{5} \binom{10}{x} 4^{x} = 320,208$ trials for developing each CER. In an approach of this type why would negative exponents be used?

Page 48. What is recommended on this page that a good PERT-Cost program wouldn't provide?

General Comments: A number of times THE procedures they have developed for the derivation of CERs were referred to. It would be helpful if these were known in order that the work being pursued at Texas A&M might be used to an advantage.

It is difficult to imagine that they are able to "massage" the data and keep everything on a proprietary basis. This cry has been used in the past to camouflage the lack of data or progress in the use of data. Therefore, it is strongly recommended that thorough documentation of all raw data and subsequent alterations of that data be required in order that revision and refinement of the cost work can be carried on after the current contracts end.

In the selection of technical variables, it was not obvious that various technical constraints were being considered in deriving the CERs.

In conclusion, what is really planned to ascertain the merits of the CER's derived?

This review of the presentation is part of the work being performed for NASA/MSC/LRP by Texas A&M under research grant NGR 44-001-027.

To: Long Range Planning Office/Manned Spacecraft Center/NASA

From: Industrial Engineering Department, Texas A&M University

Subject: Logical and Mathematical Checkout of Cost Estimating Relationships

Reference: October 19 meeting at Texas A&M

Attendees: Don Wagnor, Aubin Ferguson, Ron Konkel of LRP/MSC/NASA; A. W. Wortham and G. D. Self of Texas A&M.

The general purpose of this meeting was to discuss the procedures necessary for a cursory examination of cost estimating relationships. The suggestions contained within this report are directed toward using some simple mathematical procedures to assist in the checkout of the range of the variables contained in the CER s on a logical basis.

Since specific examples of the CER s were not available for trial computation and demonstration of the methods discussed, it will be assumed that the CER s will be similar to those developed in the Launch Vehicle Cost Model. An example of one of those equations is as follows:

$$y = 0.226x_1^{0.055} x_2^{0.0302} (100x_3)^{1.50} \qquad \text{(A)}$$

The adjustment factors for future use were unspecified and consequently were deleted.

Cost of production on liquid air frames were presented as:

$$y = 0.226\, x_1^{0.055} x_2^{.302} (100x_3)^{1.50} \quad . \qquad (1)$$

By taking the partial derivatives of cost with respect to each variable one obtains

$$\frac{dy}{dx_1} = 0.226(0.055)x_1^{-0.945} x_2^{0.302} (100x_3)^{1.50} \qquad (2)$$

$$\frac{dy}{dx_2} = 0.226(0.302)x_2^{-0.698} x_1^{0.055} (100x_3)^{1.50} \qquad (3)$$

$$\frac{dy}{dx_3} = 0.226(150)x_3^{0.50} x_1^{0.055} x_2^{0.302} \tag{4}$$

The definition and range of these variables are as follow: (B)

| | | |
|---|---|---|
| $x_1$ | stage dry weight | 3500 to 75000 lbs. |
| $x_2$ | stage volume | 2200 to 70000 ft.$^3$ |
| $x_3$ | stage mass fraction | 0.89 to 0.95 |

These partial derivatives represent the slope of the tangent for the variable which the derivative was taken with respect to and fixed values of the other two variables, or it represents the rate of change of the cost with respect to the differentiated variable.

Examining (2) it can be seen that a change in $x_1$ has almost no effect on the rate at which the cost changes; therefore, an initial selection of the range on $x_1$ could be $3,000 \leq x_1 \leq 100,000$. Since this interpretation of the slope of the line contained in the $x_1$ y plane is dependent upon the values of $x_2$ and $x_3$, their ranges should be evaluated. By inspection of $x_3$ first, since the theoretical limit is less than 1.0, one could assume a maximum value of $(100x_3)^{1.5}$ as 1,000 which does not change $\frac{dy}{dx_1}$ significantly since the maximum value of $(0.055)x_1^{-0.945}$ at the lower bound of $x_1 = 3000$ is 0.000028 and at the upper bound of $x_1 = 100,000$ it is equal to 0.00009. The minimum value of the term $(100x_3)^{1.5}$ would be approximately 800. Therefore, it is possible to look only at the range of values for $x_2$. If $x_2$ were 100,000, $x_2^{.302}$ would only be 32.36 and $\frac{dy}{dx_1}$ is approximately .21 which would be significantly different from zero if it were possible from an engineering standpoint. An engineering analysis would permit one to specify a minimum density (lbs/ft.$^3$) of say 0.2 in which case $\frac{\dot{dy}}{dx_1}$ = .03 as a maximum value for the range shown above with the constraint of density $\geq 0.2$. This would indicate that vehicles with low densities, i.e. high volume to weight ratios would expect the cost to be affected more by the weight variable than those vehicles which make up the historical data. (All vehicles had #/ft.$^3$ $\geq 1.0$.)

This evaluation process should be carried out for each of the other partial derivatives and a logical evaluation of the range of variables performed. It is suggested that the range of variables from the previous evaluation be used for each successive evaluation. This will automatically provide the final set of bounds for each of the variables involved when the last partial derivative has been evaluated.

With the partial derivatives, the total differential can be formed to obtain an expression of the slope of a hyperplane that is tangent to a hypersurface and represents the rate of change of the cost for corresponding changes in the variables or:

$$dy' = \frac{dy}{dx_1} dx_1' + \frac{dy}{dx_2} dx_2' + \frac{dy}{dx_3} dx_3'$$

From this the variance of the cost can be shown to be

$$y^2 = \frac{dy}{dx_1}^2 \sigma x_1^2 + \frac{dy}{dx_2}^2 \sigma x_2^2 + \frac{dy}{dx_3}^2 \sigma x_3^2$$

This relationship can then be used to evaluate the seriousness of an error in the CER, by testing the possible variation at each bound. These computations appear tedious from the above; however, these could be programmed for computer solution after the partial derivatives had been computed. By selecting the initial range of the variables at the desired level and using the above procedure, the best meaningful range of variables will be selected. In addition to the recommendations outlined above as a simple check on CER s, it is also suggested that each CER be accompanied by a simple correlation matrix which will reflect the normal contribution or physical interaction of one variable with another as it relates to cost. This could be placed in the form of an array which would not be symmetric. If the creation of such an array does not automatically bring about a ranking of the variables involved, it is requested that such a ranking be performed. This is to rank the most important cost variable first and so on until all variables have been ranked within a CER. It should be pointed out that under abnormal conditions any of the lesser variables of this ranking could become a cost dominating influence.

The above is a short term approach to the checkout of existing CER s. As part of this research grant work, a report on the methodology of development will be presented.

## SUMMARY

The above discussion is summarized in outline form below for each CER.

1. Form all partial derivatives
2. Program computer for evaluation
3. Compute values of the partial derivative and select range where behavior is logical based on varying one variable at a time
4. Select bounds for all variables in CER
5. Form the total differential.
6. Form the variance relation
7. Evaluate the variance relation at each bound
8. Form a simple correlation matrix for CER variables.
9. Rank variables in CER

REFERENCES

(A) Launch Vehicle Systems Cost Model, Volume II, FZM-4154, 15 June 1964, General Dynamics/Fort Worth

(B) Launch Vehicle Systems Cost Model, Progress Report 2, FZM-4070, 27 February 1964, General Dynamics/Fort Worth

To: Dr. A. W. Wortham

From: G. D. Self

Subject: Booz-Allen Applied Research Approach to Cost Estimating Relationship Development

Reference: Trip to MSC for work session on November 3, 1965.

Attendees: Messieurs Briggs, Mandell, Wagner, Ferguson, Hannaford, and Konkel of MSC; Locke and Clancy of Booz-Allen; Brents of GD/FW; and Self of Texas A&M.

The general purpose of this meeting was to discuss the specific approach that Booz-Allen has developed for use in the estimation of the parameters associated with the cost estimation relationships (CER s). Mr. Clancy was in charge of the discussion and had a few hand drawn charts to indicate the sequence of steps used to select the variables and the selection of the constants to be used in association with those variables which are combined to form the cost estimating relationship.

Their approach to the selection of variables as an exhaustive set of 'causal' variables appeared reasonable from an engineering standpoint and represented a relatively good start. Causal is used here as opposed to concomitant in that it is assumed that the variable has some functional or direct relationship to the cost of the system involved. In an effort to determine variables which represented the cost of the system over its entire life cycle, there were four general areas selected which were assumed to represent the major contributions to cost. These were combined with the thirteen subsystems to provide a cross-classification of the various blocks of costs that should be logically accounted for in the CER s. This seemed to form a reasonable base for the definition of variables.

The next major area of endeavor for the derivation of CER s was the determination of functional dependency among the variables of a specific system and the deletion of variables from consideration. This attempt was made in order to bring the number of parameters to be estimated to within the limitations placed upon the parameter estimation process by the data available. In performing this operation, two things of importance were noted; (1) it provided a graphical method for determining whether the combination of variables into a single variable was applicable for the case being considered, and (2) it tended to counteract some of the original intent of the cross-classification scheme for the selection of the variables. These two points will not be elaborated upon at this time since later in this report there are some observations which supersede any remarks that would be made here.

Assuming that the deletion and combination of variables provides the variables to be considered for the CER, the next step is to use a limited search of suitable exponents for those variables to be used in a product form for the CER. This method assumes that the CER will be of the form $y = x_1^{\alpha} x_2^{\beta} x_3^{\gamma}$. Their method uses the computer to provide an output which forms all combinations of 1/2,1, and 2 as the exponents of the variables in the CER which was used to compute the costs for all the historical data points available. The computer results are manually examined to pick the 'best' set of exponents for the CER. This is repeated with various sets of variables until the best combination of variables and exponents are found. After this is accomplished, the actual data are used with the selected variables and the resultant set of simultaneous equations are solved to obtain the new exponents for the CER. In each case the number of variables, the number of parameters to be estimated, and the number of exponents were numerically equivalent.

The approach outlined above was presented for those cases which had three data points or more. There was not a specific set of procedures set forth for those cases with less data. Booz-Allen indicated some interest in the method of combining two independent estimates. This information was provided along with a short explanation of the combination of independent estimates weighted inversely proportional to their variances. The problem of the limited data case still exists under this procedure, due to the estimates of the variance of the independent predictors which must be available and serve as the basis for combining the estimates.

At this point, the meeting turned to the subject of what is being done at Texas A&M and how will it help in the methods for the development of the CER s. A short oral presentation of the research being conducted as part of the cost research grant was made. It was generally agreeded that Booz-Allen would need more short term methods in order to accomplish their objectives within the time limitation of their contract. The suggestions relative to "Logical and Mathematical Checkout of Cost Estimating Relationships" were reiterated with particular emphasis given to the ranking and the estimation of the correlation coefficients among variables.

In general it was concluded that the methods discussed by Booz-Allen could be accomplished much quicker by a weighted regression analysis. For the single data point case, it is anticipated that Booz - Allen will devise a method for obtaining a CER and report on it at a later date.

# P A R T VI

The material contained in this section of the report represents a logical discussion of the completed work of one of the members of the research group that is working on this grant. The subject in this study is the learning curve and the discussions are in terms of the classical approach to the learning curve with some excellent examples of its application in costing and decision making. There is one point that is touched upon and which appears to be a relatively important with respect to NASA projects. This is the influence of unions and management directives upon the learning that is observed in development type projects. It should be noted that these influences could possibly reduce or camouflage the learning. However, this learning is continually being advertised as taking place in the space program in spite of the level of development effort that is being carried on. The reports in the cost contract areas continues to support the hypothesis that the learning exists, in which case the material presented in this section is timely and should provide background information in the assessment of the mistakes that could be made in improperly applying the use of learning curves to various costing situations.

This section of the report represents some of the concept that is being used in the research grant in that from time to time during the course of this grant one of the current topics that is observed in the MSC work will be abstracted and dealt with academically in order to provide the background information that should accompany that topic.

# THE LEARNING CURVE

*By*

*Grady L. Haynes*

The problem of forecasting of manpower requirements, time of production, cost per unit produced, etc., i.e., the basic parameters of industrial production, has troubled management since the introduction of modern assembly line techniques. The predictions have usually been based on personal judgments and subjective probabilities by the upper echelons of management, and have involved much time and effort.

The learning curve, often called the improvement curve, a new concept which arose from the aircraft industry during World War II, has proven to be a valuable tool in the solution of forecasting problems. The learning curve concept relies on the proven fact that a worker learns as he works; and the more the operation is repeated, the more efficient the worker becomes, with the result that the direct labor input which he contributes to each unit declines. The rate at which his skill develops, often referred to as his "percentage of learning," is fairly constant throughout, no matter how many units a worker may produce and, as we shall see, this makes possible the prediction of the total labor input for a specified number of units [1].

Probably the first notice of this "learning" took place in the Consolidated-Vultee Aircraft Corporation, San Diego, California. The company noticed that it took roughly ten times as long to complete the first two planes as it did the last two planes in a production order of 1,000 planes. Part of this decreased time was credited to improved jigs and fixtures and conveyors as the new job got under way. But much of it was due to an acquired drilling and riveting skill which comes not only with eye-hand coordination but with *muscle conditioning*. An interesting illustration of this phenomenon took place during the early part of the war.

> There was considerable variation in drilling time between hand (electric) drillers in a large airplane factory — the variation being greater than 100% from the fastest time to the slowest. An examination of drill points proved they were all in good condition. But it was found that the electric drills themselves ranged in speed from 2000 rpm to 4500 rpm. This surely was the answer. But was it? The man with the lowest time had the slowest drill and the operator who took the longest time had the fastest drill! Further investigation disclosed the foreman was an experienced operator who really "leaned" on his drill, whereas the latter, a new operator, merely steadied his drill [5].

The aircraft industry had also noticed that decreases in production times occurred with increases in numbers of units produced from their cost of production figures. Since the cost of an airframe is much higher than most mass-produced articles, any change in unit cost was easily seen. This decline meant not only lower production costs on a unit basis, but also lower production times, which was most welcome to a production-conscious wartime aircraft industry.

The process was noticed to repeat itself. Whenever a new type of airframe was introduced, the direct labor input per unit soared for the first few units, then would settle down into a gradually downward sloping curve. The curve was always very accurately predictable, and after the spread of the use of the curve from Consolidated-Vultee to the other aircraft manufacturers, the companies began to use the curves on a regular basis for cost and labor predictions [1].

The original work on learning curves was done by E. K. Yost, at the time an employee of Consolidated-Vultee Aircraft Corporation. The remarkable results achieved by Consolidated-Vultee drew the attention of the Army Air Force and the Navy Bureau of Aeronautics, and from these agencies the use of learning curves spread to other aircraft manufacturers. The concept was somewhat confined to the aircraft industry until after the war, when its use was made available to all industry through publications of the men originally responsible for its development [2].

The savings in direct labor input between certain levels of production were always in the same easily predicted proportion. It was found that the fourth unit required about 80% of the labor of the second, the tenth about 80% of the fifth, and the 2,000$^{th}$ about 80% of the 1,000$^{th}$. This led to the description of the curve as being an "80% learning curve," one which has found acceptance in nearly all forms of production.

There are three formulas which make possible the prediction of (1) cumulative average man hours for any number of units, (2) unit man hours for a specific unit, and (3) total man hours required to build a predetermined number of units. These formulas are:

(1) $Y = KX^N$, where

$Y$ = cumulative average man hours for any number of units;

$K$ = number of man hours to build first unit;

$X$ = any number of completed units;

$N = \ell n$ (% of learning curve) $\div \ell n2$.

(2) $U = (N + 1)KX^N$, where:

$U$ = unit man hours for a specific unit, and all other symbols the same as in Formula (1).

(3) $T = KX^{N+1}$, where:

$T$ = total man hours required to build a predetermined number of units and all other symbols the same as in Formula (1) [1].

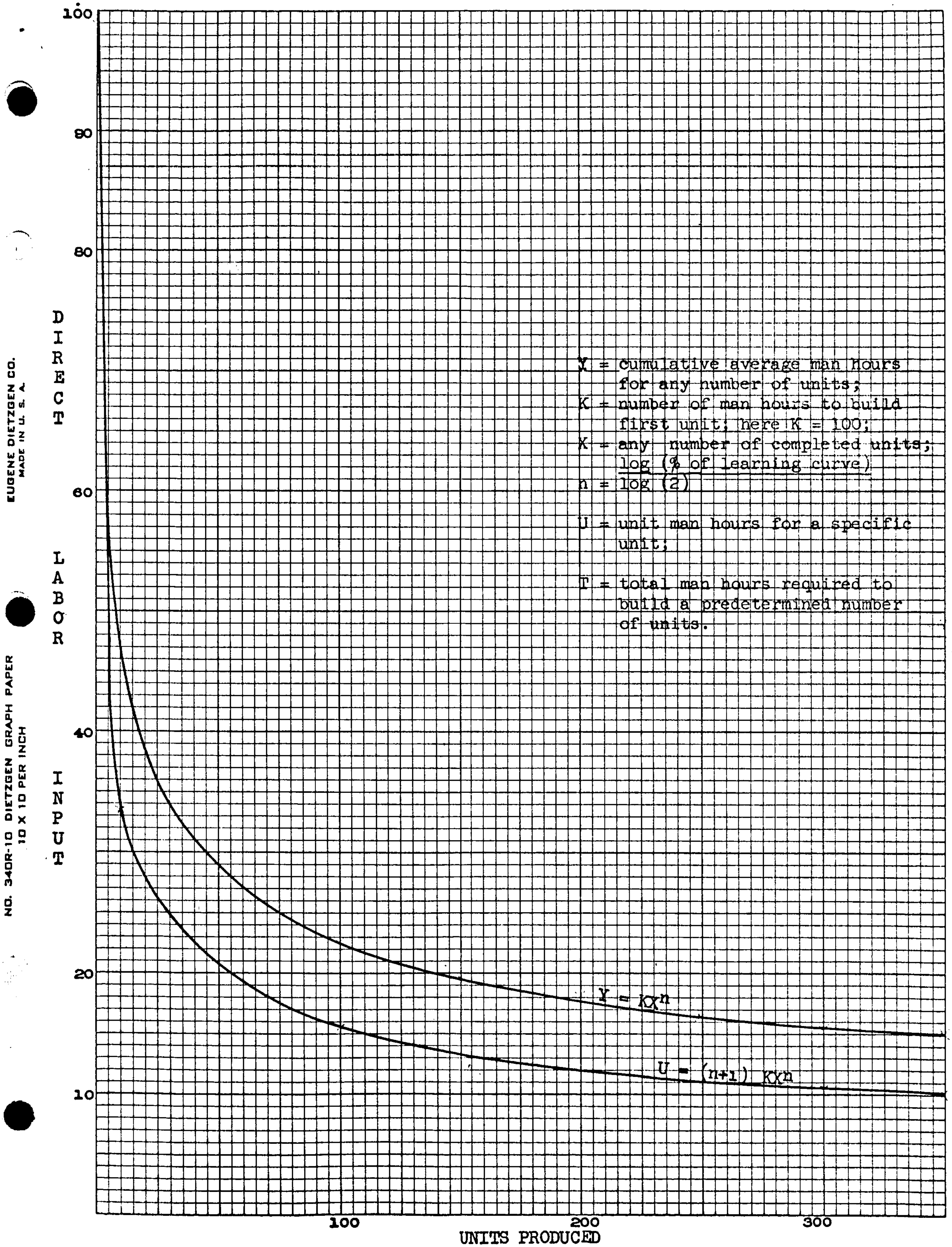
The three curves produced are shown in Figures (1), (2), and (3). The first is plotted on paper of arithmetic scale, the second on semilog paper, and the third on log-log paper. The arithmetically-scaled curves show more dramatically the changes in these three measures of production as the number of units produced increases.

Based upon continuous repetition without change in product or method of assembly, the learning curve tends to be very smooth. In actual practice, learning curves are usually erratic. This results from factors outside of actual learning such as working conditions and even the personalities of the workers themselves. New methods introduced to the assembly line cause great jumps in the curve, which are usually followed by a resumption by the curve of its normal downward slope [4].

The type of work involved also plays a part in the graph. The normal percentage of learning for manual labor assembly line production is in the 75% — 80% range, with this percentage increasing as the amount of mechanical aid to the worker increases [4]. While empirical data seems to substantiate this change in percentages, some claim that it is not the machine but the worker who learns, and thus there should be no change [3]. However, the figures most widely used are 80% and 90% for nonmechanized and mechanized labor, respectively.

There are many possible uses for the learning curve concept and also many misuses. One of the popular methods of forecasting the labor per unit produced at some future point in production is to wait until a certain number have been produced, then take the unit time of the last unit produced and connect the coordinate of this point with the coordinate of the first unit produced, and use an extension of this line

Fig. 1


DIRECT LABOR INPUT
100
90
80
60
40
20
10
UNITS PRODUCED
100
200
300
Y = cumulative average man hours for any number of units;
K = number of man hours to build first unit; here K = 100;
X = any number of completed units;
n = log (% of learning curve) / log (2)
U = unit man hours for a specific unit;
T = total man hours required to build a predetermined number of units.
Y = KX^n
U = (n+1) KX^n
EUGENE DIETZGEN CO. MADE IN U. S. A.
NO. 340R-10 DIETZGEN GRAPH PAPER 10 X 10 PER INCH

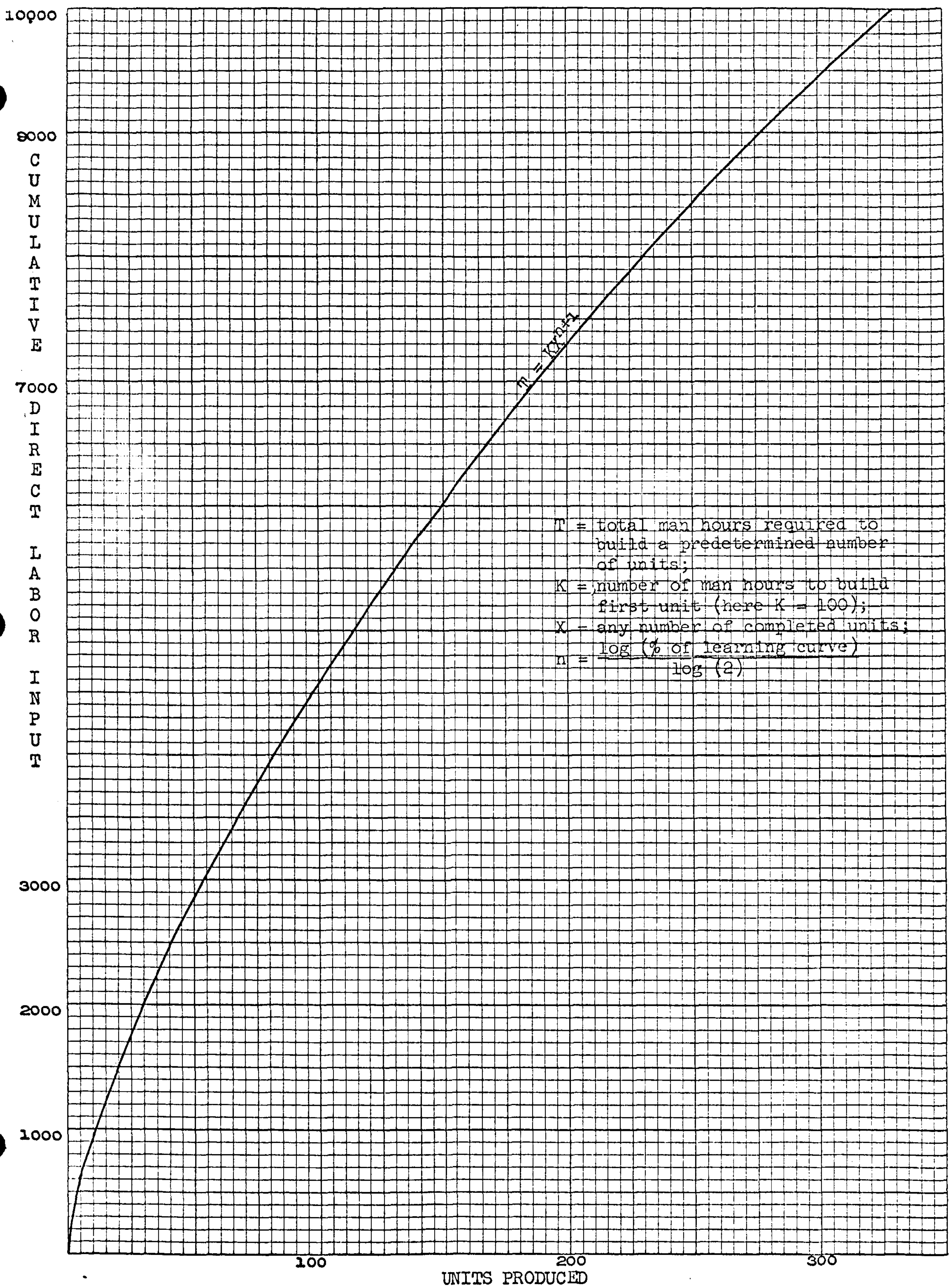
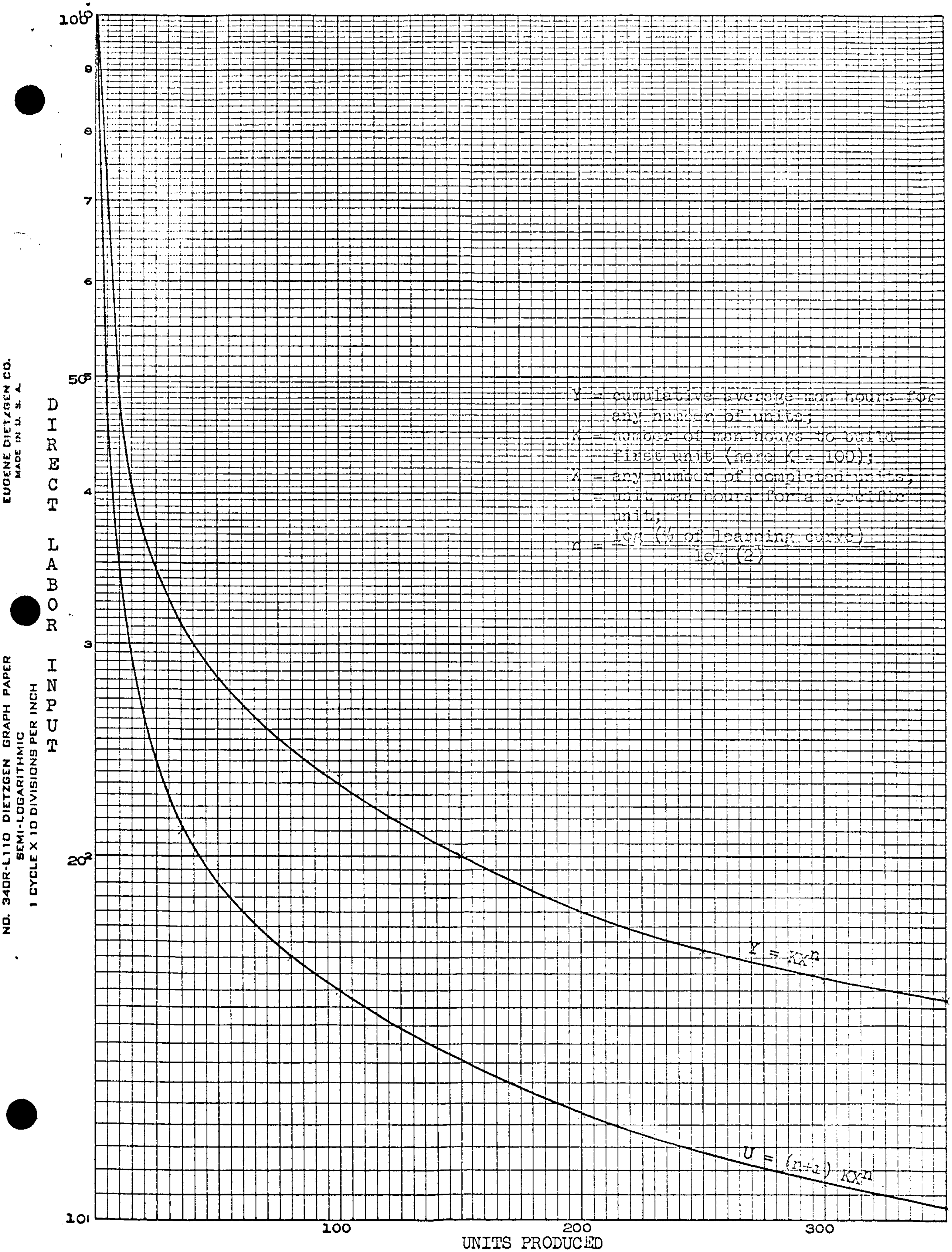
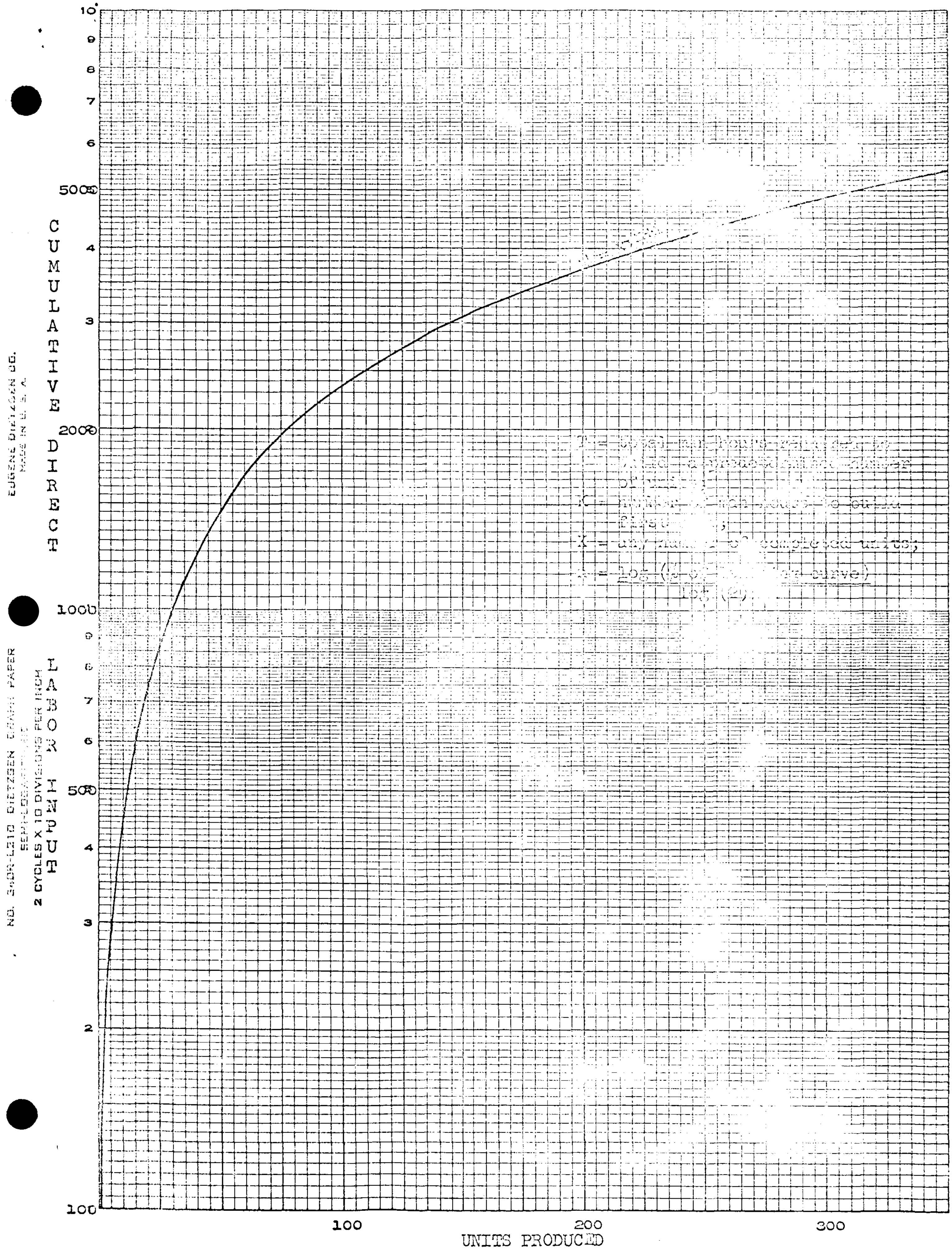

10000
9000
7000
3000
2000
1000
CUMULATIVE DIRECT LABOR INPUT
100
200
300
UNITS PRODUCED
$T = KX^{n+1}$
T = total man hours required to build a predetermined number of units;
K = number of man hours to build first unit (here K = 100);
X = any number of completed units;
$n = \frac{\log (\% \text{ of learning curve})}{\log (2)}$

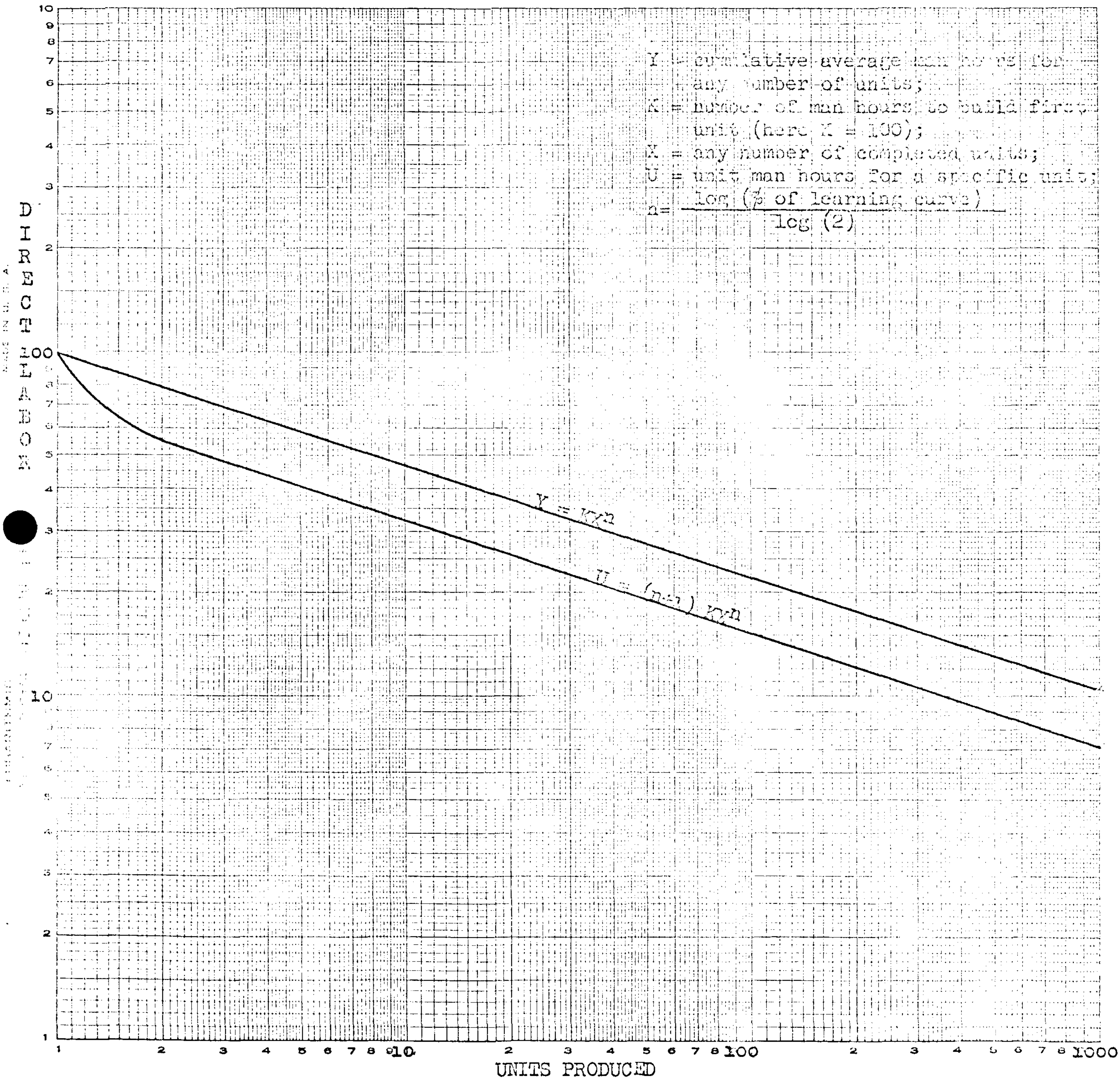
Fig. 2

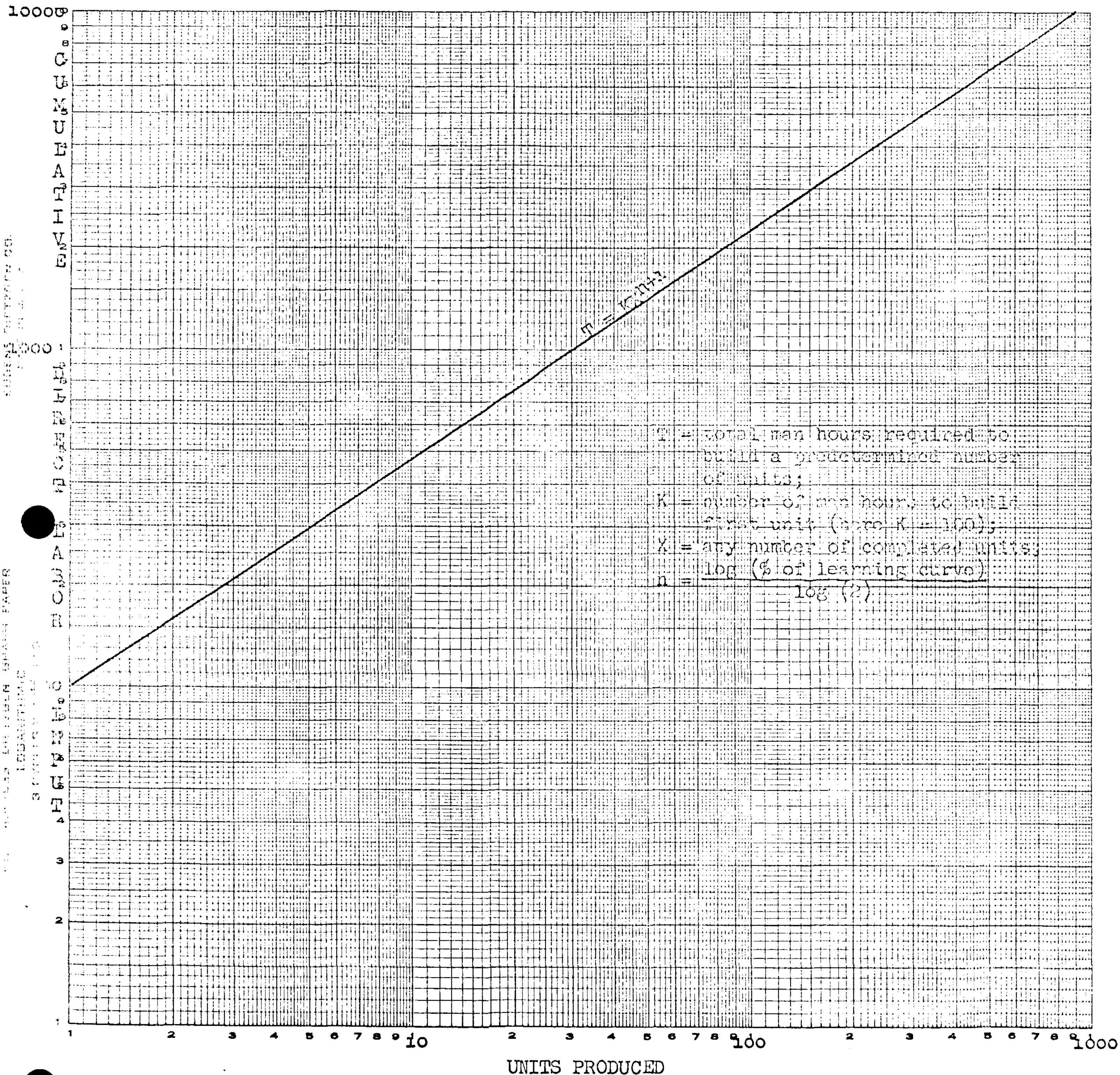


DIRECT LABOR INPUT
UNITS PRODUCED
100
200
300
10
20
50
100
Y = cumulative average man hours for any number of units;
K = number of man hours to build first unit (here K = 100);
X = any number of completed units;
U = unit man hours for a specific unit;
n = log (% of learning curve) / log (2)
Y = KX^n
U = (n+1) KX^n
NO. 340R-L110 DIETZGEN GRAPH PAPER SEMI-LOGARITHMIC 1 CYCLE X 10 DIVISIONS PER INCH
EUGENE DIETZGEN CO. MADE IN U. S. A.

Fig. 2a


CUMULATIVE DIRECT LABOR INPUT
UNITS PRODUCED
100
200
300
100
2
3
4
500
6
7
8
9
1000
2000
3
4
5000
6
7
8
9
10
completed units;
log (2)
EUGENE DIETZGEN CO.
MADE IN U. S. A.
2 CYCLES X 10 DIVISIONS PER INCH


Y = cumulative average man hours for any number of units;
K = number of man hours to build first unit (here K = 100);
X = any number of completed units;
U = unit man hours for a specific unit;
n = log (% of learning curve) / log (2)
Y = KX^n
U = (n+1) KX^n
DIRECT LABOR
UNITS PRODUCED
1
2
3
4
5
6
7
8
9
10
100
1000


10000
1000
10
1
CUMULATIVE
$T = KX^{n+1}$
T = total man hours required to build a predetermined number of units;
K = number of man hours to build first unit (here K = 100);
X = any number of completed units;
n = $\frac{\log(\% \text{ of learning curve})}{\log(2)}$
1
10
100
1000
UNITS PRODUCED

for future predictions. The danger here is choosing the points before the actual curve has had a chance to settle down to its normal downward slope. This can lead to exceedingly low predictions.

Many manufacturers have fallen into the trap of neglecting to include the direct labor input of units which are purchased from an outside supplier. For instance, a worker who uncrates an aircraft power-plant and installs it in an airframe may himself contribute eight hours of labor input, but on a curve of total labor input for the aircraft the labor involved in the production of the engine must be included. There is no net saving involved when a part is being bought from an out-side supplier since the same labor is being performed in the plant of the supplier [1].

By moving a newly-trained, highly skilled crew to an assembly line the direct labor input per unit may drop, but higher wages paid to a highly skilled crew may actually result in increased costs. Similarly, new engineering or tooling may result in lower production times but higher costs. The method which produces the lowest labor input per unit may not always be the optimum method of production.

Care must be taken to justify claims that drops in direct labor input per unit are the result of "learning." In some instances shrewd management and good personnel handling can bring about drops in labor input without any "learning" having occurred [1].

For prediction of future labor times before the first unit has been produced, a combination of learning curves with time and motion studies has proved highly reliable. It has been shown that for most mass-produced articles involving large quantities in production reliable estimates of labor times from time studies should be around the 1,000$^{th}$ unit.

From Formula (2) it can be shown that the slope of the direct labor input per unit-produced curve will approach zero for large values of X, indicating that a point is reached where the decrease in labor input with each succeeding unit is negligible. Thus it is possible to predict labor input for any unit by plotting the optimum labor time, derived by the time study method, at the $1{,}000^{th}$ unit on log-log paper and drawing a straight line of slope [log (% of learning)/log 2] back to the ordinate. After production is under way and the time for the first unit is precisely know, the graph can be revised with the line again drawn from the $1{,}000^{th}$ unit optimum time point, but through the actual time of the first unit [3].

Illustrations in three specific areas (pricing, make or buy situations and manpower requirement predictions) will make the uses of the learning curve concept more meaningful.

*The Webster Machine Company.* Following the outbreak of the Korean War the Army's own facilities were inadequate to meet the demand for gun barrels. Accordingly, the work was "farmed out" to private companies, one of which was the Webster Machine Company.

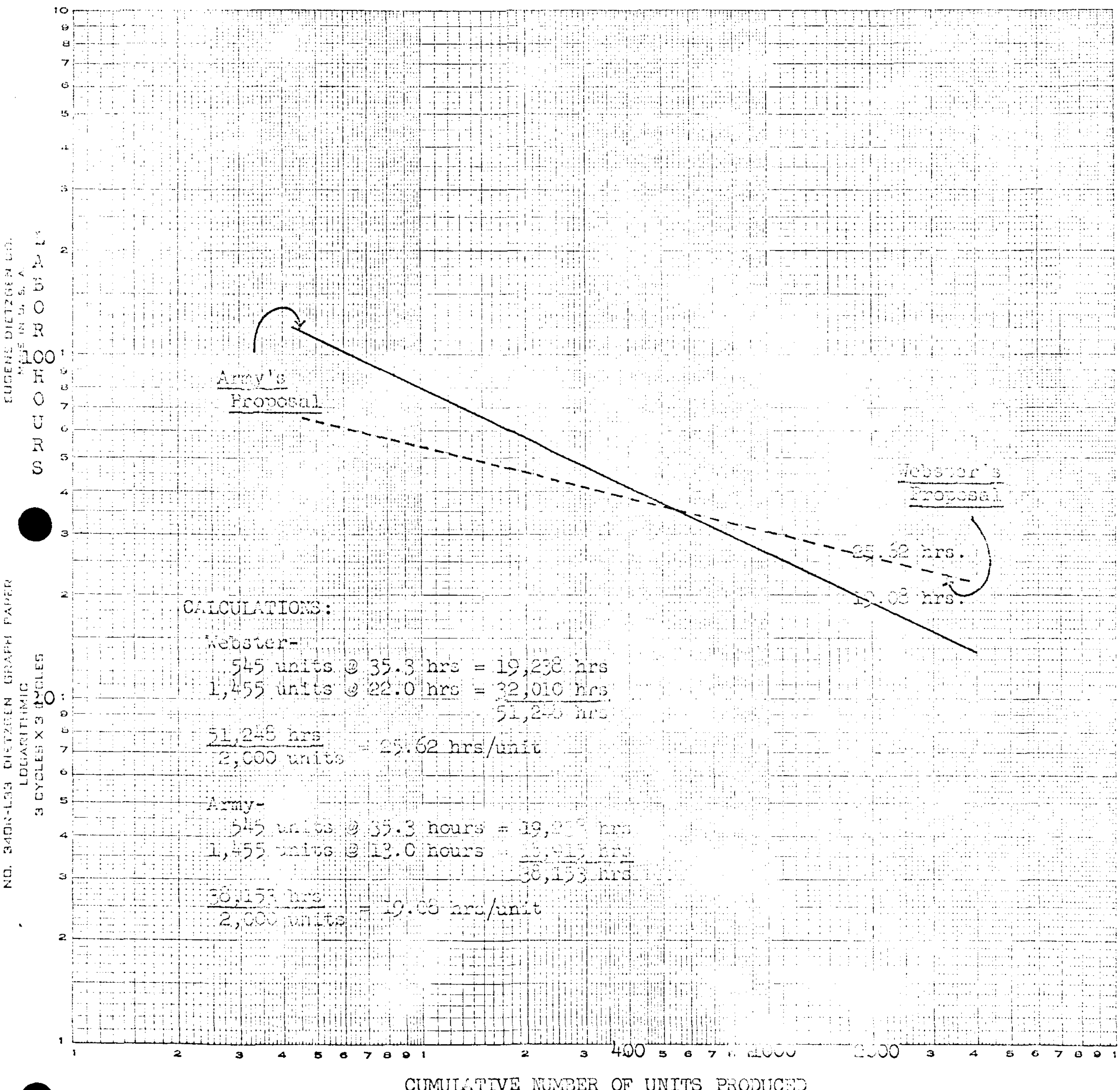
The contract called for a single rough forming operation on 2,000 gun barrels; in other words, machine time (rather than assembly time) represented a large part of the work. The initial contract price was $76.70 per unit subject to price redetermination. When the price was redetermined, Webster had completed 545 units at an average direct labor cost, verified by audit and time study, of 35.3 hours per unit. What direct labor cost should be charged in computing the price for the uncompleted unit? Both parties recognized the fact that the 35.3 hour figure was too high for the contract as a whole and that because of some sort of learning factor or increase in the contractor's "know-how," the new figure should be lower.

Webster argued for 22 hours per unit for the uncompleted units, pointing out that the Army had seven years' experience in this operation whereas Webster had never done it before. The figure of 35.3 hours was recognized as the cumulative average for the first 545 units and, accordingly, was plotted as one plot on the graph. To get the second point, which was necessary to construct the curve, the company computed the cumulative average for all 2,000 units (including the 545 produced to date) on the basis of 22 hours per unit for the uncompleted units and thus arrived at an overall average of 25.62 hours. A straight line drawn between the points had a slope, or rate of learning, of 85% between doubled quantities. This curve (Figure 4) showed that the direct labor theoretically required for the last unit would be 19.6 hours.

The Army supply officer took a very different view. He argued that since the Army was performing the operation itself in 11.6 hours, 13 hours was a fair figure for the uncompleted units. (Webster did not possess the Army's special hoists.) His proposal (Figure 3) was computed in the same manner as Webster's. The slope of his line was a phenomenal 71.5%. In order to obtain the cumulative average of 19.08 hours for all 2,000 units called for by his proposal, Webster would have had to produce the last unit in 9.82 hours. This was 1.78 hours less than the time the Army was taking to produce the gun barrels, and the Army had seven years' more experience on the job.

Clearly, Webster's proposal was the more reasonable, especially in view of the fact that machine time represented the greater part of the work, which meant a lower rate of learning (or higher percentage of

Fig. 4



MAN HOURS
Army's Proposal
Webster's Proposal
25.62 hrs.
19.08 hrs.
CALCULATIONS:
Webster-
545 units @ 35.3 hrs = 19,238 hrs
1,455 units @ 22.0 hrs = 32,010 hrs
51,248 hrs
51,248 hrs / 2,000 units = 25.62 hrs/unit
Army-
545 units @ 35.3 hours = 19,238 hrs
1,455 units @ 13.0 hours = 18,915 hrs
38,153 hrs
38,153 hrs / 2,000 units = 19.08 hrs/unit
100
10
400
1000
2000
CUMULATIVE NUMBER OF UNITS PRODUCED
EUGENE DIETZGEN CO. MADE IN U.S.A.
NO. 340R-L33 DIETZGEN GRAPH PAPER LOGARITHMIC 3 CYCLES X 3 CYCLES

labor times required between doubled quantities). The parties finally agreed on an 85% curve [1].

Make or buy decisions concern the choice of whether to make an item or procure it from an outside manufacturer.

*Lee Aircraft Company.* Early in 1952 the Lee Aircraft Company was faced with a cutback in its production as a result of the Air Force stretch-out program. Consequently, it was inclined to cancel some of its subcontracts and pull the work back into its own shop to keep it fully occupied.

One subcontract which it thought of canceling was with the Roberts Manufacturing Company for 372 landing flap assemblies — an item which it also was manufacturing in its own plant. To arrive at a comparison of its own and Roberts' costs of manufacturing the assemblies, Lee decided to plot the respective learning curves.

Lee had already produced 165 assemblies, with a figure of 445 hours for the 165$^{th}$ unit, and was well along the downward slope of its learning curve; continuation of the curve indicated a total labor input of 111,000 hours for 372 additional units. In comparison, The Roberts Company, while apparently a more efficient producer of the item, was just getting started on its learning curve. If it went on, it would be able to produce the 165$^{th}$ unit at an expenditure of 402 hours — 43 hours less than Lee — but continuation of its curve from the earlier, higher point *at which it then was* indicated a total labor input of 164,000 hours for the 372 units, or 53,000 more than Lee.

The foregoing analysis served to pinpoint the question for management's judgment. In the short run, it was more economical for Lee to cancel the subcontract and do the work itself. In the long run, however, it would be less expensive to leave the work with Roberts inasmuch as it could produce the landing flaps for about 10% less labor once it had gotten as far out on its learning curve as Lee. Therefore, the decision hinged largely on the probable total future demand for landing flaps of this type. Since this total future demand was difficult to measure, Lee decided to take advantage of the direct labor savings offered at the time, which amounted to over $300,000. Accordingly, it canceled the contract with Roberts [1].

Learning curves can play a most vital role in manpower planning, as illustrated by the following example.

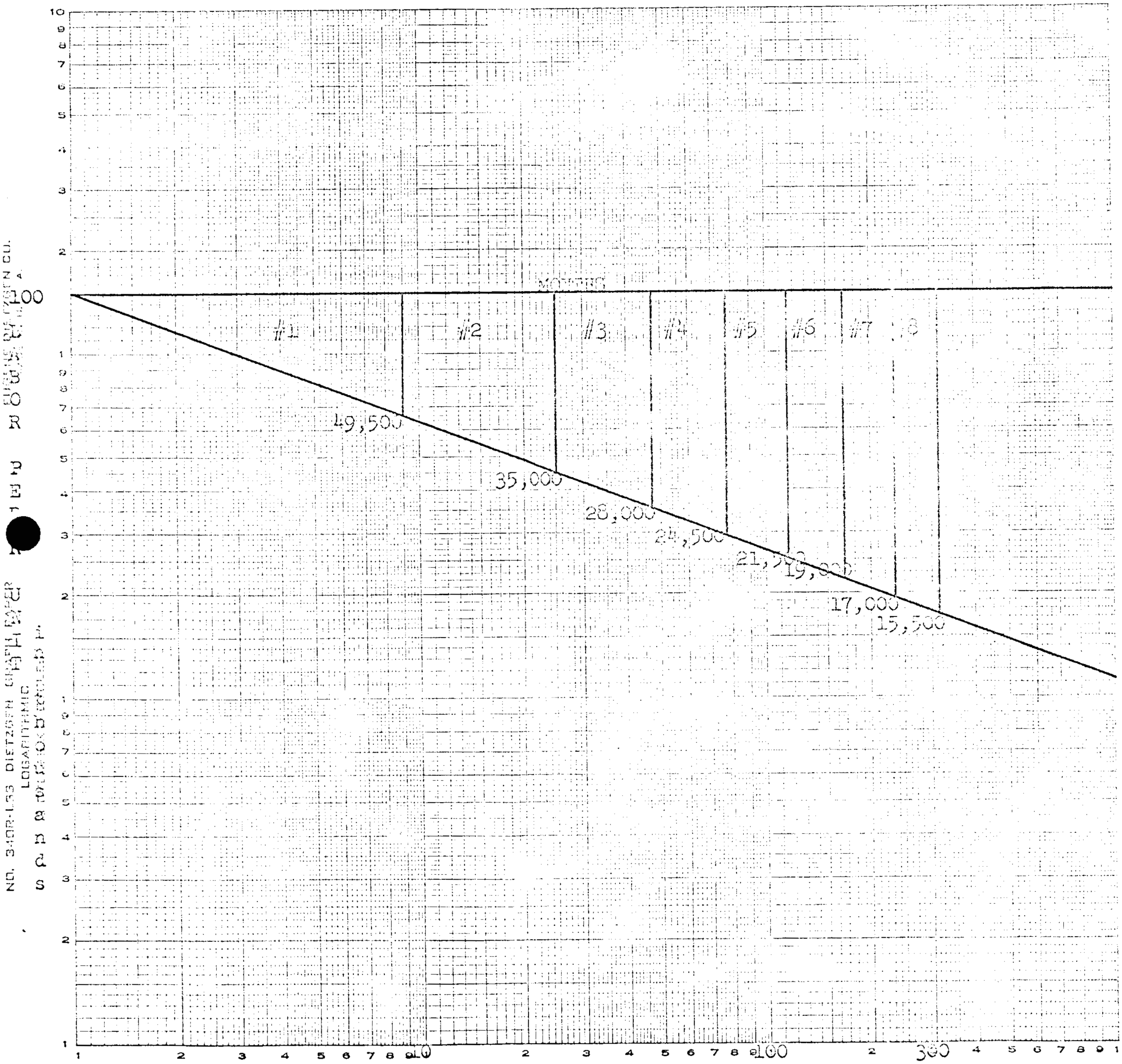
A manufacturing company determines that the demand for its new self-propelled diesel railroad car calls for a production schedule like that in the three lefthand columns of Table I, after allowing time for tooling, engineering, and procurement. Its first step would be to work out a learning curve for these units, with the production schedule representing the point of departure and the size of the work force the dependent variable.

A rate of learning of 80% between doubled quantities produced a learning curve shown in Figure (5). The monthly delivery schedule is indicated by the vertical lines; the points where they intersect the "curve" represent the cumulative averages of direct labor hour requirements as of the end of each month.

By taking these averages from Figure (5) and multiplying them by the cumulative number of units produced, the cumulative total hours for *all* preceding months can be obtained (fourth and fifth columns of Table I).

Table I. Calculation of Manpower Requirements

| Month | Units per Month | Cumulative Number of Units | Cumulative Averages (from Fig. 4) | Cumulative Total Hours | Total Hours per Month | Total Direct Employees per Month |
|---|---|---|---|---|---|---|
| 1 | 9 | 9 | 49,500 | 445,500 | 445,500 | 2,228 |
| 2 | 16 | 25 | 35,000 | 875,000 | 429,500 | 2,148 |
| 3 | 22 | 47 | 28,000 | 1,316,000 | 441,000 | 2,205 |
| 4 | 30 | 77 | 24,500 | 1,886,500 | 570,500 | 2,853 |
| 5 | 40 | 117 | 21,500 | 2,515,500 | 629,000 | 3,145 |
| 6 | 52 | 169 | 19,000 | 3,211,000 | 695,000 | 3,478 |
| 7 | 65 | 234 | 17,000 | 3,978,000 | 767,000 | 3,835 |
| 8 | 80 | 314 | 15,500 | 4,867,000 | 889,000 | 4,445 |




Cumulative Number of Units Produced

If the latter figures are divided by the number of hours worked per month (in this case 200), the needed number of direct employees can be obtained. Thus, it can be seen from Table I (rightmost column) that about 2,228 employees are required in the first month, and about double that number in the last month, whereas the number of units produced per month increases about nine times.

This same procedure could be used to determine the manpower requirements for each manpower classification [1].

From the results which have thus far been obtained through the use of the learning curve, it is rather surprising that the use of this technique has not been more extensive, and that it has not received more attention in statistical and quality control literature [2].

# Bibliography

1. Andress, Frank J., "Learning Curve as a Production Tool," *Harvard Business Review* (Jan.-Feb., 1954), pp. 87-97.

2. Bennett, C. A., L. W. Smith and W. F. Stevenson, "The Application of a Learning Curve to a Maintenance Problem," Presented at the Second Annual Quality Control Symposium of the Dallas-Fort Worth Section, American Society for Quality Control, March 16, 1957.

3. Dennington, C. T., "Practical Learning Curves Improve Shop Performance," *Iron Age* (November 26, 1953), pp. 108-111.

4. Hadley, J. R., "Learning Curves on Log-Log Paper," *Advanced Management* (April, 1950), pp. 16-17.

5. Nissley, Harold R., "The Importance of Learning Curves in Setting Job Shop Standards," *Mill and Factory* (May, 1949), pp. 119-121.

# P A R T VII

This section of the report is intended to provide a relatively complete bibliography of the literature that has been published in the last decade on the subject areas that are being considered as a part of this research grant. It is anticipated that the bibliography will be extended with each additional report during the remainder of the term of the grant. The literature listed in this section is being reviewed by the members of the research group in the areas of their particular interest and research areas.

# BIBLIOGRAPHY


1. Atichison, J. and Silvey, S. D., "Maximum - Likelihood Estimation of Parameters Subject to Restraints," Annals of Mathematical Statistics, 29 (1958), pages 813-828.

2. Aitken, A. C., "On Least Squares and Linear Combinations of Observations,"Proceedings of the Royal Society of Edinburgh, 55 (1935), pages 48-8.

3. Alchiam, Armen, "Reliability of Progress Curves in Air-frame Production," Econometrica Vol. 31, No. 4, October, 1963, page 679.

4. Anderson, T. W., "The Choice of the Degree of a Polynomial Regression as a Multiple Decision Problem," Annals of Mathematical Statistics, 33 (1962), pages 255-265.

5. Babbar, M. M., "Distributions of Solutions of a set of Linear Equations (with an application to linear programming)," JASA, 50 (1955), pages 854-69.

6. Beale, E. M. L., "On minimizing a convex function subject to linear inequalities," Journal of the Royal Statistical Society, Series B, 17 (1955), pages 173-184.

7. Blischke, W. R., "Estimating the Parameters of Mixtures of Binomial Distributions," JASA, 59, No. 306, 1964, pages 510-528.

8. Boot, J. C. G., "Binding Constraint Procedures of Quadratic Programming," Econometrica, Vol. 31, #3, July 1963,page 464.

9. Box, G. E. P., "The Exploration and Exploitation of Response Surfaces: Some General Consideration and Examples," Biometrics, 10 (1954), pages 16-60.

10. Box, G. E. P., and Draper, Norman R., "A Basis for the Selection of a Response Surface Design," JASA, 54 (1959), pages 622-654.

11. Box, G. E. P. and Hunter, J. S., "A Confidence Region for the Solution of a Set of Simultaneous Equations with an Application to Experimental Design," Biometrics, 41 (1954), pages 190-9.

12. Box, G. E. P. and Lucas, H. L., "Design of Experiments in Non-Linear Situations," Statistical Techniques Research Group (Princeton University), Technical Report No. 15, submitted for publication in Biometrics.

13. Box, G. E. P. and Wilson, R. B., "On the Experimental Attainment of Optimum Conditions," Journal of the Royal Statistical Society, Series B, 13 (1951), pages 1-45.

14. Brunk, H. D., "On the Estimation of Parameters Restricted by Inequalities," Annals of Mathematical Statistics, 29 (1958), pages 437-454.

15. Carroll, Charles W., "The Created Response Surface Technique for Optimizing Non-linear Restrained Systems," ORSA, Vol. 5 No. 4, pages 551-554.

16. Charnes, A. and Cooper, W. W., "Deterministic Equivalents for Optimizing and Satisfizing Under Chance Constraints," ORSO, Vol. 11, No. 1, pages 18-39.

17. Charnes, A. & Lemke, C.E., "Minimization of Non-linear Seperable Convex Functionals," Naval Research Logistics Quarterly, 1:301-312, 1954.

18. Cohen, A. C., "Estimating of logarithmic - normal distribution by maximum likelihood," JASA, 46 (1951), pages 206-12.

19. Cornell, Richard G., "A method for fitting linear combination of exponentials," Biometrics, 18 (1962), pages 104-113.

20. Cornell, Richard G., "A New Estimation Procedure for Linear Combinations of Exponentials," Oak Ridge National Laboratory Report ORNL 2120, Oak Ridge, Tennessee, 1956.

21. Crow, Edwin L., "Generality of Confidence Inervals for a Regression Function," JASA, 50 (1955), pages 850-853.

22. Cunia, T., "Least Squares Estimates and Parabolic Regression with Restricted Location for the Stationary Point," JASA, 59 (1964), No. 306, pages 564-571.

23. Cunia, T., "Weighted least squares method and construction of volume tables," Forest Science, 10 (June, 1964).

24. Daniels, H. E., "A Distribution-Free Test for Regression Parameters," Annals of Mathematical Statistics, 25 (1954), pages 499-513.

25. Daniels, H. E., "The Theory of Position Finding," Journal of Royal Statistical Society, (B), 13 (1951), pages 186-207.

26. Dantzig, George B., "Linear Programming under uncertainty," Management Sciences, 1 (1955), pages 197-206.

27. Dantzig, George B., "Maximization of a linear function of variables subject to linear inequalities," in Koopmans, T. C., Editor, Activity Analysis of Production and Allocation, New York: John Wiley and Sons, 1951, pages 339-47.

28. Dantzig, George B., "On the significance of solving linear programming problems with some integer variables," Econometrica, 28: 30-44, 1960.

29. Durbin, J., "A Note on Regression When There is Extraneous Information about one of the Coefficients," JASA, 48 (1953), pages 799-808.

30. Fields, R. I., Krumer, C. Y., and Clunies-Ross, C. W., "Joint Estimation of the Parameters of Two Normal Populations," JASA, 57 (1962), pages 446-454.

31. Fisher, W. D., "A Note on Curve Fitting with Minimum Deviation by Linear Programming," JASA, 56 (1961), pages 359-362.

32. Folks, J. L., "Comparison of Designs for Exploration of Response Relationships," paper read at Chicago at the 18th Annual Meeting, American Statistical Association, December, 1958.

33. Forgenson, Dale W., "Multiple Regression Analysis of a Poisson Process," JASA, 56 (1961), pages 235-245.

34. Fox, Peter D., "A Theory of Cost Effectiveness for Military Systems Analysis," ORSA, Vol. 13, No. 2, pages 191-201.

35. Geisser, S., "Bayesian Estimation in Multivariate Analysis," U. S. Army Mathematics Research Center, University of Wisconsin, Report No. 503, August, 1964.

36. Geory, R. C., "Non-Linear Functional Relationship Between Two Variables When one Variable is Controlled," JASA, 48 (1953), pages 94-103.

37. Greenberg, B. G. and Sarhan, A. E., "Matrix Inversion, Its Interest and Application in Analysis of Data," JASA, 54 (1959), pages 755-765.

38. Gumbel, E. J., "Bivariate Exponential Distributions," JASA, 55 (1961), pages 698-707.

39. Hardy, G. H., Littlewood, J. E., and Polya, G., Inequalities, Cambridge University Press, 2nd Edition, 1952.

40. Hartley, H. O., "Exact Confidence Regions for the Parameters in Non-Linear Regression Laws," Biometrika, 51(3&4), 1964, pages 347-53.

41. Hartley, H. O., "Modified Gauss Newton Method for the fitting of Non-linear Regression Functions," Technometrics, 3 (1961), pages 269-280.

42. Hartley, H. O., "Non-linear Programming by the Simplex Method," Econometrica, Vol. 29, No. 2, April, 1961, page: 223.

43. Hartley, H. O., "Non-linear Programming for seperable objective functions and constraints," (Dittoed), 1959.

44. Hartley, H. O., "The Analysis and Design of Experiment with the Help of Computers," -

45. Hartley, H. O., "The Application of Mathematical Programming to Statistical Least Squares Estimation with Contsraints," Technometrics, No. 1, NSF Grant 14236, 1960.

46. Hartley, H. O., and Broemeling, T. D., "Near Optimal Properties of Confidence Region for the Parameters in Non-Linear Regression Laws," Unpublished Report, Texas A&M University, 1965.

47. Hartley, H. O., and Booker, Aaron, "Non-linear Least Squares Estimation," Annals of Mathematical Statistics, 36 (1965), pages 638-650.

48. Hartley, H. O., and Hocking, R., "Convex programming by tangential approximation," Vol. 9, #4, Management Science, pages 600-613, 1963.

49. Hendricks, J. W., "A System for Development of Mathematical Models," U.S. Naval Radiological Defense Laboratory, 1964.

50. Hildreth, Clifford, "Point Estimates of Ordinates of Concave Functions," JASA, 49 (1954), pages 598-619.

51. Hill, Bruce M., "Information for estimating the proportions in mixtures of exponential and normal distributions," JASA, 58 (1963), pages 918-932.

52. Hill, Bruce M., "The Three-Parameters Lognormal Distribution and Bayesian Analysis of a Point-Source Epidemics," JASA, 58 (1963), pages 72-84.

53. Hill, L. S., "Towards a Coordinated Approach to the Structuring of Cost and Design Models: A Case Example," ORSA, Vol. 13, No. 4, pages 619-631.

54. Hock, Irving, "Estimation of Production Function Parameters Combining Time-Series and Cross - Section Data," Econometrica, Vol. 30, No. 1, January, 1962, page 34.

55. Hoel, P. G., "Confidence Bands for Polynomial Curves," Annals of Mathematical Statistics, 25 (1954), pages 534-542.

56. Hoel, P. G., "Minimax Designs in Two Dimesional Regression," Annals of Mathematical Statistics, 36 (1965), pages 271-294.

57. Halperin, M., "Confidence Interval Estimation in Non-Linear Regression," Program 360, Applied Mathematics Department SRRC-RR-62-78, Sperry Rand Research Center.

58. Halperin, M., "Interval Estimation of Non-Linear Parametric Functions, II, " JASA, 59 (1964), No. 305, pages 168-181.

59. Halperin, M., and Mantel, Nathan, "Interval Estimation of Non-linear Parametric Function," JASA, 58 (1963), pages 611-27.

60. Hotelling, H., "The Experimental Determination of the Maximum of a Function," Annals of Mathematical Statistics, 12 (1941), pages 20-45.

61. Houthakker, H. S., "The Capacity Method of Quadratic Programming," Econometrica, Vol. 28, No. 1, January, 1960, page 62.

62. Huzurbayar, V. S., "The likelihood equation consistency and the maxima of the likelihood function," Annals Eugenics, 14 (1948), pages 185-198.

63. Kale, B. K., "On the solution of Likelihood Equations by iteration processes. The multiparametric case," Biometrika, 49 (1962), pages 479-86.

64. Karst, O. J., "Linear curve fitting using least deviation," JASA, 53 (1958), pages 118-32.

65. Katy, Morris W., "Admissible and Minimax Estimates of Parameters in Truncated Spaces," Annâls of Mathematical Statistics, 34 (1963), pages 462-479.

66. Kiefer, J. and Wolfowitiz, J., "Optimum Designs in Regression Problems," Annals of Mathematical Statistics, 30 (1965), pages 271-294.

67. Kimball, B. F., "A Multiple Group Least Squares' Problem and the Significance of the Associated Orthogonal Polynomials," JASA, 48 (1953), pages 320-335.

68. Land, A. H., Doig, A. G., "An Automatic Method of Solving Discrete Programming Problems," ORSA, Vol. 28, No. 3, July, 1960, page 497.

70. Larson, Harold J. and Bancroft, T. A., "Sequential Model Building for Prediction in Regression Analysis," Annals of Mathematical Statistics, 34 (1963), pages 462-479.

71. Lentner, M. M. and Boehler, R. J., "Some Inferences about Gamma Parameters with an Application to a Reliability Problem," JASA, 58 (1963), pages 670-677.

72. Lewis, Henry R.,"The Data Enrichment Method," ORSA, Vol. 5, No. 4, pages 551-554.

73. Lewish, W. T., "Linear Estimation in Convex Parameter Spaces," Ph. D. Thesis, Iowa State, 1963.

74. Mason, David M., "Functional models and experimental designs for characterizing response curves and surfaces," from Boum, E. L., Heady,Earl O., and Blackmore, John, Editors, Methodological Procedures in the Economic Analysis of Fertilkerg Use Data, Ames, Iowa: Iowa State College Press, 1956.

75. Papert, S., "Some Mathematical Models of Learning," Information Theory, The Fourth London Symposium, (Edited by Colin Cherry), Butterworth, London and Washington, 1961.

76. Rider, Paul R., "Estimating the parameters of mixed Poisson, binomial, and Weibull distribution by the method of moments," Bulletin of the International Statistical Institute, 39, Part 2 (1962), pages 225-32.

77. Rider, Paul R., "The method of moments applied to a mixture of two exponential distributions," Annals of Mathematical Statistics, 32 (1961), pages 163-7.

78. Robbins, Herbert, "Asymptotic Subminimax Solutions of Compound Statistical Decision Problems," Proc. Second Berkeley Symp. Math. Statistical Problem, University of California Press, 1951, pages 131-148.

79. Sacks, Jerome, "Generalized Bayes Solutions in Estimation Problems," Annals of Mathematical Statistics, 34 (1963), pages 751-768.

80. Sumuel, Ester, "Note on Estimating Ordered Parameters," Annals of Mathematical Statistics, 36 (1965), pages 698-702.

81. Savage, L. J.,"The Subjective basis of Statistical Practice," Manuscript, University of Michigan, 1961.

82. Shetty, C. M., "A Simplified Procedure for Quadratic Programming," Econometrica, Vol. 11, #2, pages 248-260.

83. Singleton, R. R., "A method for minimizing the sum of absolute values of deviations," Annals of Mathematical Statistics," 11 (1940), pages 301-10.

84. Skibinsky, M. and Cote, L., "On the Inadmissibility of Some Standard Estimates in the Presence of Prior Information," Annals of Mathematical Statistics, 34 (1963), pages 539-548.

85. Speckman, Janace A. and Cornell, Richard G., "Estimation for A One-Parameter Exponential Model," JASA, Vol. 60, No. 310, June, 1965, pages 560-572.

86. Stone, M., "The Opinion Pool," Annals of Mathematical Statistics," 32, 1961, pages 1339-1342.

87. Theil, H., "On the Use of Incomplete Prior Information in Regression Analysis," JASA, 58, 1963, pages 401-14.

88. Tiao, G. C. and Zellner, A., "Bayes's Theorem and the Use of Prior Knowledge in Regression Analysis," Biometrika, 44, 1964a, pages 315-27.

89. Tiao, G. C., and Zellner, A., "On the Bayesian Estimation of Multivariate Regression," Journal of the Royal Statistics Society, Series B, 1964b, pages 277-85.

90. Tiao, G. C. and Zellner, A., "Bayes' Theorems and the Use of Prior Knowledge in Regression Analysis," Technometrics, Report, No. 14, Department of Statistics, University of Wisconsin, 1963.

91. Takayama, T. and Judge, G. G., "Spatial Equilibrium and Quadratic Programming," Journal of Farm Economics, Vol. 46, No. 1, February, 1964, pages 67-93.

92. Tukey, J. W., "The Experimental Approach in the Teaching of Statistics Use of Experiments in Teaching Engineering Statistics," Unsolved Problems of Experimental Statistics, JASA, 49 (1954), No. 268.

93. Turner, M. E., Monroe, K. J., and Lucas, H. L., "Generalized Asymptotic Regression and Non-Linear Path Analysis," Biometrics, 17, pages 120-43, 1961.

94. Wagner, Harvey M., "Linear Programming techniques for regression analysis," Journal of the American Statistical Association, 54: 206-212, 1959.

95. Wagner, Harvey M., "On the Distribution of Solutions in Linear Programming Problems," JASA, 53 (1958), pages 161-163.

96. Walsh, John E., "Use of Linearized Non-linear Regression for Simulations Involving Monte Carlo," ORSA, Vol. 11, #2, pages 228-235.

97. Walters, A. A., "Production and Cost Functions," Econometrica, Vol. 31, No. 1-2, January-April, 1963.

98. Weiss, Lionel, "On Estimating Scale and Location Parameters," JASA, 58 (1963), pages 658-59.

99. Williams, E. J., "Exact Fiducial Limits in Non-Linear Estimation," Journal of Royal Statistic Society, B, 24, pages 125-39.

100. Wolf, Philip, "The simplex method for quadratic programming," Econometrics, 27: 382-355, 1959.

101. Wolf, Philip, "The RAND Symposium on Mathematical Programming," Editor, R-351, The RAND Corp., 1960.

102. Zellner, Arnold and Chetty, V. Koruppan, "Prediction and Decision Problems in Regression from the Bayesian Point of View," JASA, Vol. 60, No. 310, June, 1965, pages 608-616.

103. Zoutendijk, G., "Maximizing a function in a convex region," Journal of the Royal Statistical Society, Series B, 21: 338-355, 1959.